融合内容与关系的学术社交媒体上跨学科用户推荐模型研究*

吴小兰[1] 章成志[2]

[1] 南京师范大学新闻与传播学院网络与新媒体系 南京 210046

[2] 南京理工大学经济管理学院信息管理系 南京 210094

摘 要：[目的/意义]在学术社交媒体快速发展的今天，开展跨学科研究或者寻求跨学科合作时，很多科研合作起始于社交媒体上的相识或关注，因此开展社交媒体上跨学科用户推荐非常有意义。社交媒体上主要存在“媒体”(代表内容)、“社交”(代表关系)两大类数据，因此本文开展了融合内容与关系的社交媒体跨学科用户推荐。[方法/过程]在基于向量空间模型的用户表示之后，本文借助用户内容信息计算用户领域专业度，根据关系数据测度用户跨学科距离，同时结合用户关系网络 PageRank 值给出推荐结果。[结果/结论]以科学网为例，实现“图书情报”“计算机”“新闻与传媒”“高等教育”“生物学”这5个领域内的跨学科用户推荐，并经人工实验测试检验，表明推荐结果在一定程度上能满足推荐需求。

关键词：跨学科用户 推荐模型 跨学科距离 学术社交媒体

分类号：G203

DOI：10.13266/j.issn.0252-3116.2020.09.011

1 引言

在日常生活中，人们如果遇到一些问题，常希望得到网上相关领域用户的解答。在现实中，同样有很多场景例证这种需求[1-2]，如项目评审中需要领域用户(或专家)对项目评阅、审核时，需要系统根据项目相关性推荐用户。随着 Internet 技术与 Web2.0 技术的发展，网络在科研中的重要性也越来越强，F. Barjak 发现学者们在使用社交媒体和网络站点文献时，经常引用他们与对方沟通的观点[3]；G. Kirkup 通过访谈发现博客成为了一种新的学术生产方式，同时较便捷地实现了博主与评论者的对话及思想碰撞[4]。在此背景下，面向科研人员的学术社交媒体应运而生[5]。所谓学术社交媒体，又称为“学术社交网络”(Academic Social Networking，后简称 ASN)，是指能够帮助研究者与其他学者建立社会网络和促进科研活动的一种在线服务、工具或者是平台[6]。目前国外使用较多的学术社交媒体有 ResearchGate、Academia.edu、Mendeley，国内的 ScholarMate(科研之友)、科学网博客、学术圈等也很受学者欢迎。

随着学术社交媒体的进一步发展，很多不同领域的学者、团队、机构等纷纷加入并借助学术社交媒体平台来分享自己的成果，实现与同行交流，这为跨学科研究与合作提供了有益帮助[7]，但也为学术社交媒体上寻找到合适的跨学科用户(或专家)增加了难度。因此，在学术社交媒体上开展跨学科用户推荐非常重要，也非常有意义，这便是本文研究动机——学术社交媒体上跨学科用户推荐。该推荐可满足用户在一些场景下对跨领域用户的寻找，帮助他们准确地找到相关的专家(通常，如果一个用户对某一事物精通或者说有独到的见解，我们称他(她)是这一领域的专家)。

与一般信息检索系统不同，推荐系统并不直接提供问题答案，只是提供解决问题的途径，因此该推荐模型与现有专家推荐系统有很多相似之处，本文致力于构建这样一个跨学科用户的推荐模型：融合“社交”数据(代表关系)与“媒体”数据(代表内容)，以实现学术

* 本文系国家社会科学青年基金项目“社交媒体视域下的跨学科用户发现及其推荐研究”(项目编号：17CTQ047)研究成果之一。

作者简介：吴小兰(ORCID：0000-0003-1869-1738)，副教授，博士；章成志(ORCID：0000-0001-9522-2914)，教授，博士，博士生导师，通讯作者，E-mail：zhangcz@njust.edu.cn。

收稿日期：2019-10-09 修回日期：2020-02-01 本文起止页码：95-103 本文责任编辑：易飞

社交媒体上的跨学科用户推荐。

2 学术社交媒体及其跨学科研究现状

作为社交媒体中面向学术科研的一类，学术社交媒体的内容主要来自于科研用户，因此学术社交媒体上的研究更多地集中在以下 3 个方面：①用户使用行为与学科差异性方面，如 H. Meishar-Tal 等根据使用与满足理论分析了以色列 3 个学术机构 298 名用户的问卷，发现用户使用 ASN 的主要目的是消费信息，分享信息、与他人互动相对较弱[8]；J. L. Ortega 通过分析 6 132 名拥有学术社交媒体账号的西班牙国家研究理事会成员的 Profiles，发现 Academia. edu 上有大量的人文科学家和社会科学家，而 ResearchGate 比较受生物学家欢迎[9]，在每个 ASN 上都可以观察到学科差异；A. M. Elsayed 分析收集到的 315 名阿拉伯科研人员问卷，发现拥有 ResearchGate 账号人员较多且多数来自自然科学和应用科学领域[10]。②学术影响方面，如 J. Priem 等认为社交网络在科学计量学 2.0 中显示出潜力，Social Webmetrics 丰富了传统文献引用的评价指标，甚至其上的资源聚合服务可以防止"睡美人"现象[11]；W. Gunn 认为 Mendeley 已成为评价学术期刊影响因子的替代计量指标之一[12]。③学术资源推荐方面，如 L. Jing 等基于合著顺序、最新合作时间和合作次数提出了基于随机游走模型的 ACRec 模型，该模型可以实现合作者推荐[13]；V. A. Rohani 等提出了一种冷启动的 ECSN 算法向用户推荐学术项目[14]。

但随着学术社交媒体的发展，学者们逐渐关注到 ASN 在跨学科研究中的作用，如 S. J. Oh 等[15]通过对 Mendeley 上 21 679 个 Group 及其上的 67 562 条关系的分析，发现 43 124 条关系（占 63.8%）发生在用户所在学科领域内，余下的 36.2% 表明用户可能突破了学科限制，加入了其他学科群组。此外，Mendeley 上 Group 表现出明显的学科多样性，为来自不同背景的研究人员提供一个平台，让他们在共同感兴趣的问题上找到彼此并进行合作，有助于促进多学科合作；J. Jiang 等[16]通过构建 Mendeley 上群－群成员耦合网络和群－群跟随耦合网络，研究了 Mendeley 上群与学科之间的交互，并指出学术社交媒体上群有望为跨学科研究提供丰富的语料；X. Wu 等[17]基于科学网的学科分类体系和科研用户所填研究方向、好友关系等数据，并借用生物学领域的亲缘物种物种均匀度指标发现了高影响力跨学科用户。

通过上面描述，可以看出学术社交媒体对资源推荐有优势，对跨学科研究也有其促进作用，但是尚且缺乏社交媒体上跨学科用户推荐方面的研究，而跨学科用户推荐对促进跨学科项目合作乃至科研创新都起着至关重要的作用。

从某种程度上来说，跨学科用户推荐与传统推荐系统有很多相似之处。目前关于用户推荐系统研究最多的是专家推荐，专家推荐的一个比较明确的定义是由 T. Reichling 等[18]提出：专家推荐是为了满足用户在特定场景下对专家的需要，帮助他们及时准确找到相关领域专家来解决问题的推荐系统。现有专家推荐系统主要有基于专家知识信息的推荐、基于专家社会网络分析的推荐以及综合两者的方法。其中，基于专家知识信息的推荐是在构建专家信息后，计算专家知识和用户需求之间的匹配度来实现，如李明等引入信息熵求解需求模型和专家知识模型的匹配度来产生推荐[19]；基于专家社会网络分析的推荐通过建立专家社会网络关系提取出专家关联度（关联度体现在专家之间的学术问题或群体交流等潜在交流中），然后利用网络关系挖掘出专家，如 J. M. Kleinberg[20] 提出了基于图的专家知识推荐方法。综合两者的方法有许云红[21]集结社会网络分析和语义分析的专家知识推荐，H. Kautz 等[22]的 Referral Web。单纯地基于专家信息进行推荐会存在冷启动问题，单纯地基于关系角度挖掘会缺乏语义理解，因此本文我们提出融合内容（即"媒体"数据）与关系（即"社交"数据）的跨学科用户推荐，以帮助用户更方便地寻找到他们感兴趣的具有一定影响力的跨学科用户。

3 研究思路与关键技术指标

清华大学彭兰教授[23]认为社交媒体有两个主要特征：一是内容生成与社交相结合；二是平台上主角是用户，而不是网站运营者。因此本文提出将内容与社交融合一起进行跨学科用户推荐的方法，其研究思路见图 1 。

通过图 1，可以看出，该推荐模型主要包括下面 3 个部分：

（1）用户知识表示模型。文本挖掘中文档较多地采用向量空间模型（Vector Space Model，VSM）来表示，这样文档就会以向量形式定义到实数域中，使得自然语言具有可计算性。在本文中，推荐需求和跨学科用户都看成文档，其 VSM 构造过程见 3.1 节。

（2）推荐相似度计算。本文从领域方向和领域知识两个角度进行相似度计算。如果相似度值高，说明

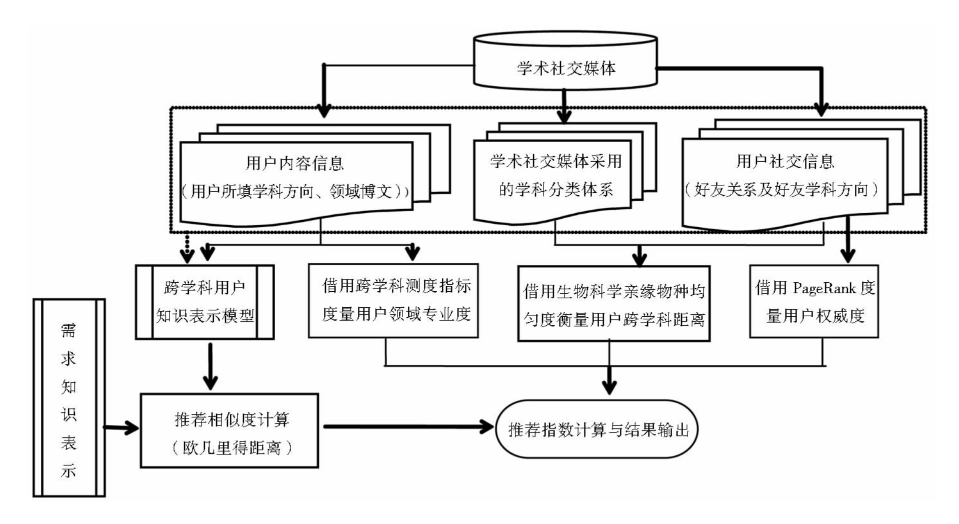


图 1　融合内容与关系的跨学科用户推荐模型

推荐需求与用户信息匹配程度高，反之匹配程度低，详细技术见 3.2 节。

(3)推荐指数计算及结果输出。得到领域相似度值后，对相似度超过指定阈值的用户，结合用户领域专业度、跨学科距离、学者 PageRank 值给出推荐顺序，其详细技术见 3.3 节。

3.1　用户知识表示模型

为了计算用户信息和需求信息之间的距离，我们采用了 VSM 来定义用户模型，该模型基本思想是假定文档中词出现的概率在内容和位置上是相互独立的。向量空间模型由 G. Salton 等于 1974 年提出[24]，后来被应用到 Web Watcher、Fab 等很多个性化推荐系统中，并且使用效果较好。

本文对用户博文分词、去除停用词后构成用户文档向量。令领域知识中包含的关键词集合为 T，且 $T=\{T_1,T_2,\cdots,T_n\}$，总数为 n。用户集合标记为 E，且 $E=\{E_1,E_2,\cdots,E_m\}$，则用户向量表示为：$E_i=<(T_1,w_{E_1}),(T_2,w_{E_2}),\cdots,(T_n,w_{E_n})>$，其中，$w_{E_i}$ 表示关键词 T_i 的权重，权重 w 采用 TF * IDF 来计算，计算公式为：

$$w(T_j,E_i)=tf(T_j,E_i)*\log(\cdots)\frac{m+1}{m_{T_j}+1}+1)\quad 式(1)$$

式中 $tf(T_j,E_i)$ 为关键词 T_j 在用户 E_i 博文中出现的频率，m 为总用户数，m_{T_j} 为有关键词 T_j 出现的用户数。

3.2　推荐相似度计算

为了克服相似度计算的片面性，本文采用需求与领域方向、领域知识的双层相似度计算，双层相似度计算公式定义如下：

$$Sim(U_i,E_j)=\alpha Sim(D_{U_i},D_{E_j})+\beta Sim(K_{U_i},K_{E_j})$$

式(2)

其中，$\alpha+\beta=1$，且 $\alpha\leqslant\beta\leqslant1$，式中 α 体现领域方向(Direction)在推荐中的重要程度，β 体现领域知识(Knowledge)在推荐中的重要程度。$Sim(D_{U_i},D_{E_j})$ 为用户需求 U_i 和领域用户 E_j 之间的领域方向(Direction)相似度；$Sim(K_{U_i},K_{K_j})$ 为用户需求 U_i 和领域用户 E_j 之间的领域知识(Knowledge)相似度。

在计算用户之间相似度时，欧氏距离(Euclidean Distance)是比较直观、常见的一种相似度算法。欧氏距离越小，两个用户相似度就越大，欧氏距离越大，两个用户相似度就越小。在日常使用中，一般习惯于将相似度与 1 进行比较，相似度在数值上反映为 0 < = Similarity(X,Y) < = 1，越接近 1，相似度越高；那么我们在使用欧氏距离时，可以通过 1/(Distance(X,Y))来实现上一理念。假定用户需求 i 与跨学科用户 j 的文本特征表示为 $\overrightarrow{v_i}=(w_{i1},w_{i2},\cdots,w_{im})$ 和 $\overrightarrow{v_j}=(w_{j1},w_{j2},\cdots,w_{jm}$，二者之间基于欧氏距离的相似度计算公式为：

$$Sim(\overrightarrow{v_i},\overrightarrow{v_j})=\frac{1}{Distance(\overrightarrow{v_i},\overrightarrow{v_j})}=\frac{1}{\sqrt{\sum_{t=1}^{m}(w_{it}-w_{jt})^2}}$$

式(3)

3.3　推荐指数计算及跨学科用户推荐

上述相似度计算可有效地检索出领域方向与领域知识上满足需求的领域用户，但可能会出现推荐精度不高的现象，即一些跨学科能力较强的用户的推荐位置可能被靠后。在项目评审中，文献[25-26]从科学计量角度将专家科研能力纳入专家遴选标准，重点考察专家研究方向、学术水平、科研经验、知名度等信息。因此，本文也从科学计量角度去计算跨学科用户推荐指数(Recommendation Index，RI)。在本文中，我们主要从学者领域专业度、跨学科距离和学者网络中心性

角度来考虑，认为推荐指数是领域专业度（S）、跨学科距离（IDD）、学者网络 PageRank 值（PR）的加权线性组合，即：

$$RI = \alpha 1 \cdot S + \alpha 2 \cdot IDD + \alpha 3 \cdot PR \quad 式(4)$$

式（4）中，学者领域专业度值反映了用户在某一学科的影响力，跨学科距离主要反映学者跨学科程度，学者网络 PageRank 值反映用户关系网中权威度，各个指标说明依次如下：

（1）领域专业度。作者专业度（specialization）[27]是 A. L. Porter 等学者提出的，描述了在指定时间内一个研究者发表的论文分布的学科情况。该指标在杨良斌等[28]的跨学科测度方法中也被使用，其计算公式如下：

$$S_p = \frac{\sum m_i^{\,2}}{(\sum m_i)^2} \quad 式(5)$$

式（5）中，m_i 表示属于学科类别 i 的文章数，S_p 越低，说明作者跨学科研究程度越高，专业度越低；S_p 越高，说明研究越集中，其专业度越高。

但和晋飞等[29]发现 S_p 得出的值大部分都在 0.5 以下，而实际情况是作者的研究工作专业度并不太低，为此，我们采用和晋飞等提出的 S 公式：

$$S = \frac{2\left[\frac{n+1}{2} - Q\right]}{n-1} \quad 式(6)$$

其中，$Q = 1 \times f_1 + 2 \times f_2 + \cdots + n \times f_n$，$f_n$ 表示每个学科类别的论文数占总论文数的比例，且 $f_1 \geqslant f_2 \geqslant \cdots \geqslant f_n$。

（2）跨学科距离。实际上，不管是用 Sp 还是用 S 公式来计算用户专业度，其值大只能说明用户在该学科的博文所占比例较多，并不能代表用户的跨学科属性，因此我们进一步借用了跨学科距离来衡量用户的跨学科程度。

跨学科距离（interdisciplinary distance，IDD）可以用来衡量用户的跨学科程度，其思想来源于文献[30]。在文献[30]中，作者指出 IDD 思想来源于文献[31]中的生物科学亲缘物种均匀度——Phylogenetic Species Evenness，其计算公式如下：

$$PSE = \frac{m\,diag(C)'M - M'CM}{m^2 - \bar{m}_i m} \quad 式(7)$$

式（7）中，C 为物种亲缘关系树的关系矩阵，$diag(C)'$为 C 对角矩阵，M 为物种个数分布的列向量，m 为物种关系树所有物种个数，$\bar{m}_i$ 为物种关系树每个分支上平均物种个数。

通过式（7），可以看出，IDD 指标的计算是在物种亲缘关系树上来实现的，为此，本文我们创新性地利用用户好友的领域学科分布构造用户学科亲缘树。所谓学科亲缘关系树是仿照生物学领域提出的一个概念。在生物学领域，生物分类学家根据各类生物间亲缘关系的远近，把各类生物安置在有分枝树状的图表上，简明地表示生物的进化历程和亲缘关系，称为“进化树”，不同分支上的生物差异较同一分支上的差异大。实际上，在科学发展过程中，科学门类也同样发生了这样的演化并分支过程，有些学科之间自被创建以来就表现出明显的“亲缘”关系。为了突出学科间“亲缘”与“远疏”关系，本文同样借用了“学科亲缘关系树”这一概念来表达用户与学科间的相似程度。在本文实验中，科学网博客上博主的学科亲缘关系树构造过程示例如图 2 所示：

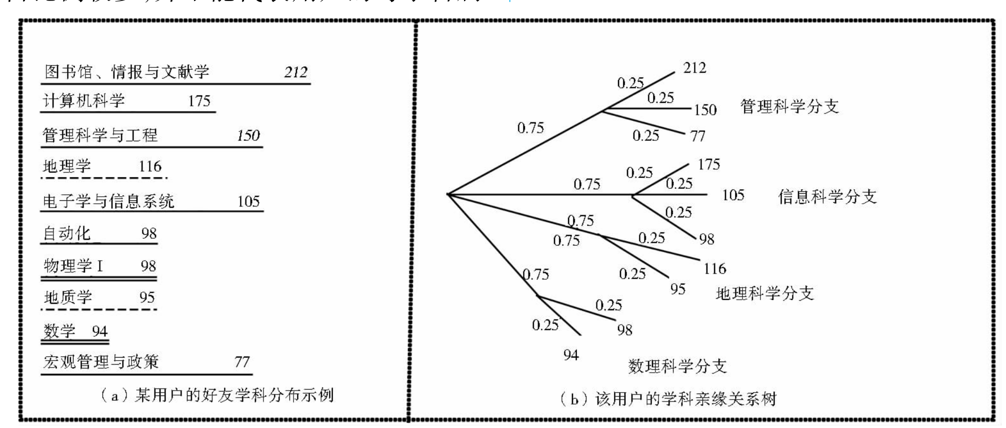


图 2 用户学科亲缘关系树构造过程示例

为了构造跨学科关系亲缘树，我们另外需要获得平台的学科分类体系，因此本文爬取了科学网博客上的学科分类体系：包含 105 个二级学科类别（后文统称为“二级学科类别”）的 8 个学科大类（后文统称为“一级学科类别”），具体类别数见表 1。

表 1　科学网博客上学科分类类目

一级学科类别名称	包含的二级学科类别数目(个)	二级学科类别示例(仅给出 3 个)
生命科学	20	微生物学、植物学、生态学
医学科学	31	呼吸系统、循环系统、消化系统
化学科学	7	无机化学、有机化学、物理化学
工程材料	9	金属材料、无机非金属材料、有机高分子材料
信息科学	5	电子学与信息系统、计算机科学、自动化
地球科学	6	地理学、地质学、地球化学
数理科学	5	数学、力学、物理学 I
管理综合	22	管理科学与工程、宏观管理与政策、图书馆、情报与文献学

在本文中，假如统计某用户好友的二级学科类别后得到的情况如图 2(a)(按二级学科类别数目降序排列，数值代表属于该类别内的好友数)。根据图 2(a)的学科分布，我们便可以构造得到如图 2(b)所示的学科亲缘关系树。其构造过程如下：①对照表 1，可以看出图 2(a)中相同下划线标出的二级学科同属于同一一级学科类目，因此该用户的好友其实是主要来自于 4 个一级学科门类，因此最终可以得到图 2(b)具有 4 个分支。②根据图 2(a)中各个二级学科类别下好友数目，便可以得到图 2(b)中对应 4 个分支下 tips(指学科亲缘关系树的枝端)的数目，如示例中该用户有 212 个好友属于“图书馆、情报与文献学”分支，那么该 tips 下数目便是 212；3)。我们认为 tips 节点到根节点(指学科亲缘关系树的根端)的长度为 1，属于同一分支的学科为同域学科，不同分支的学科为异域学科，显然同域学科的学科共性要强，因此本文我们设置属于同一个分支下的不同 tips 共有的长度为 0.75，各个 tips 到二级学科类别的长度为 0.25，从而最后得到了图 2(b)。

通常，跨学科距离越大的用户，其好友学科分布越分散、跨学科距离越小的用户，其好友学科分布越集中，因此，图 2 示例中用户的好友关系亲缘关系树构造好后，便可以得到 IDD 指标的各个值，如 $C = [c_{ij}]_{10\times10}$(其中 c_{ij} 为物种 i 与 j 到根节点的共享长度)，$M' = [116,95,212,105,150,98,175,94,98,77]$，$m = 122\ 0$，$\overline{m}_i = 122$，则该用户的 $IDD = 0.846\ 7$。

(3)学者关系网络 PageRank 值。PageRank 算法[32]是由拉里 · 佩奇和谢尔盖 · 布林等于 1998 年提出的网页排序算法，其思想为：一个网页重要，要么是有链接指向的网页多，要么是有链接指向它的网络重要。因此用户关系网中用户 PageRank 值可以反映用户权威度。为此，我们构建跨学科用户(供 371 位)及其好友关系网，并计算了每个用户的 PageRank 值，得到的本文实例数据如表 2(Top50)所示：

表 2　跨学科用户的 PageRank 样例(Top50)

用户名	PageRank 值	用户名	PageRank 值	用户名	PageRank 值
胡春松	0.013 3	王德华	0.007 1	章成志	0.005 3
罗春元	0.012 5	刘洋	0.006 9	李银生	0.005 2
陈儒军	0.012 4	倪鹏云	0.006 8	曾新林	0.005 2
孙学军	0.012 2	许培扬	0.006 8	张成岗	0.005 2
刘钢	0.012	唐常杰	0.006 7	段俊	0.005 1
赵凤光	0.010 1	李维	0.006 3	黄富强	0.005 1
武夷山	0.009 7	张欣	0.006 1	张宇	0.005 1
房松	0.009 3	高建国	0.006 1	周可真	0.005
吴锦宇	0.009 2	张志东	0.006	周涛	0.005
周公朴	0.008 4	徐明昆	0.005 9	马臻	0.005
吕喆	0.007 8	曹聪	0.005 7	闵应骅	0.005
黄成	0.007 6	曾泳春	0.005 6	李侠	0.005
王春艳	0.007 6	迟菲	0.005 6	邵立晶	0.004 9
赵美娣	0.007 6	喻海良	0.005 6	何宏	0.004 9
陈安	0.007 5	曹广福	0.005 4	刘桂锋	0.004 8
王号	0.007 4	李本先	0.005 4	刘进平	0.004 8
汪育才	0.007 2	梁进	0.005 3		

4　跨学科用户推荐模型实验与结果分析

4.1　跨学科用户推荐模型实验设计

根据邱均平等[33]的研究，图书情报学者跨学科研究的前 20 个学科领域依次是“计算机软件及计算机应用”“新闻与传媒”“高等教育”等，对照着科学网博客上的学科分类，我们最终选择了“图书情报学”“计算机应用”“新闻与传媒”“高等教育”“生物学”这 5 个学科领域的跨学科用户作为本文实验对象，即前期基于主题得到的 284 位跨学科用户[34]和基于关系角度得到的 128 位跨学科用户[17]，累计 371 位跨学科用户(有重复用户，因此最终用户数小于二者总和)。同时加入进模型的还有这些用户的累计 43 598 篇领域博文，领域博文数目如表 3 所示：

表 3　跨学科用户的领域博文数

领域名称	领域博文数(篇)
图书情报	9 143
计算机	5 044
新闻与传媒	13 169
高等教育	8 738
生物学	7 504

随后，我们根据博文所属博主 ID 对 43 598 篇博文进行了合并，生成 371 个跨学科用户文档，并计算了用户领域专业度 S。在用户文档形成后，我们统计了博

文中词的出现情况，得到 577 044 个词，经去重处理之后得到 26 727 个特征词，随后我们采用 TF-IDF 对这些特征权重进行了计算，最终得到 371 个跨学科用户文档向量，加入进数据库。

与此同时，我们提取了这些用户的好友关系网（371 个节点与 7 919 条边），计算了 PageRank 值，也根据用户好友学科分布计算了用户的跨学科距离 IDD。

基于上述数据之后，我们借助 Visual studio Code + Nodejs 架构设计了一个推荐模型，推荐过程如下：用户在推荐模型中选择领域方向或领域知识之后，返回排名 Top15 的跨学科用户列表，并同时给出推荐用户的其他信息（如工作单位与职称）浏览链接。在该模型中，跨学科用户排序的思路是：领域方向（direction）与领域知识（knowledge）相似度加权组合，其中各个权值设置如下（领域专业度值，跨学科距离，Pagerank 值都归一化之后进行权重设置）：$\alpha=\beta=0.5$，$\alpha_1=\alpha_2=\alpha_3$，推荐模型界面如图 3 所示：

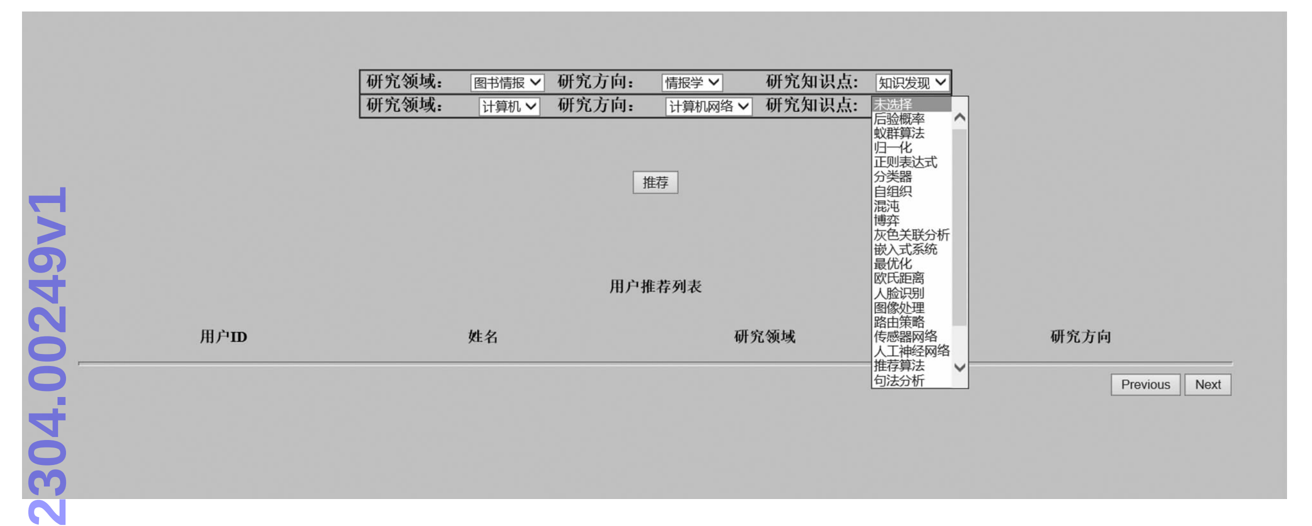


图 3　跨学科用户推荐界面

在图 3 中，我们可以根据推荐需求实现推荐，其过程如下：①用户先选择两个不同研究领域，共有“图书情报”“计算机”“新闻与传媒”“高等教育”“生物学”5 个学科领域；②随后，用户可以选择级联的“研究方向”，本推荐模型中的“研究方向”主要来自于 371 位跨学科用户填写的二级学科方向；③继用户选择“研究方向”之后，进一步可以选择级联的“研究知识点”，其中“研究知识点”来自于 371 位用户的博文特征；④最后，点击“推荐”查看推荐结果，其界面如图 4 所示：

用户 ID	姓名	研究领域	研究方向
34250	郭崇慧	管理综合->管理科学与工程	“知识发现”:0.017828
36782	章成志	管理综合->图书馆、情报与文献学->情报学	“知识发现”:0.000737118“句法分析”:0.00167824
711035	王婷	信息科学->计算机科学->自然语言理解与机器翻译	“知识发现”:0.00323566
39714	李斌	信息科学->计算机科学->自然语言理解与机器翻译	“句法分析”:0.00124909
362400	李维	信息科学	“句法分析”:0.00111121
280034	许培扬	管理综合->图书馆、情报与文献学->情报学	“知识发现”:0.0010635
5573	刘耀	信息科学->计算机科学->自然语言理解与机器翻译	
460603	韩红旗	信息科学->计算机科学->自然语言理解与机器翻译	
652078	李伟钢	信息科学->计算机科学->计算机应用技术	“知识发现”:0.0013404
39723	刘桂锋	管理综合->图书馆、情报与文献学->情报学	
220689	苏学	管理综合->图书馆、情报与文献学->情报学	

图 4　跨学科用户推荐结果

4.2　跨学科用户推荐模型结果分析

为了验证推荐质量，我们采用了人工打分方法来测评。从“图书与情报”领域找到了 10 位博士生与硕士生，从“计算机”领域与“通信”领域找到了 18 位硕士生对推荐结果进行质量评估，其测评过程如下：①首先测评用户根据自身推荐需要选择感兴趣的学科领域、学科方向和学科知识，然后点击“推荐”按钮实现推荐；②在获得到推荐之后，测评用户对推荐的 10 个

跨学科用户逐一评分，依据的评分准则是 1－5 分(越高代表该用户与推荐需求越匹配)。

在实验人员进行评分过程中，我们要求实验人员必须逐一单击"推荐应用"中"用户 ID"(具有超链接)，去查看用户的详细信息。同时，我们要求测评人员逐一记录推荐需求与推荐用户。一个测评人员至少选择一组(不超过两组)推荐需求，然后对推荐结果进行测评，最终得到 35 份推荐需求及其测评结果，推荐需求如表 4 所示：

表 4　测评人员的推荐需求统计

学科领域组合	频次	学科方向组合
(图书情报，计算机)	8	(文献学，自然语言理解与机器学习)(情报学，自然语言理解与机器学习)(情报学，自然语言理解与机器学习)(图书情报文献其他，计算机体系结构)(计算机体系结构，情报学)(信息安全，情报学)(计算机体系结构，情报学)(情报学，计算机体系结构)
(新闻与传媒，高等教育)	6	(新闻事业经营管理，教育心理学)(传播学，教育心理学)、(教育心理学，新闻史)(职业技术教育学，新闻史)(教学论，新闻事业教育管理)(新闻业务，教学论)
(生物学，新闻与传媒)	5	(人类遗传学，传播学)(遗传学研究新技术与方法，新闻史)、(新闻史，基因组学)(基因组学，新闻业务)(微生物遗传学，传播学)
(生物学，计算机)	4	(生物信息学，自然语言理解与机器学习)(生物信息学，计算机应用技术)(计算机应用技术，基因组学)(计算机软件，人类遗传学)
(图书情报，新闻与传媒)	4	(文献学，传播学)(传播学，情报学)(图书情报文献学其他，新闻事业经营管理)(情报学，新闻事业经营管理)
(高等教育，生物学)	3	(教育心理学，人类遗传学)(基因组学，教育技术学)(微生物遗传学，教育管理学)
(计算机，高等教育)	3	(计算机软件，教育信心理学)(信息安全，教学论)(计算机网络，教育技术学)
(图书情报，高等教育)	2	(情报学，教育心理学)(图书情报文献其他，教学论)

通过表 4，可以看出：①测评中"(图书情报，计算机)"推荐需求最多，这可能是因为测评人员主要来自于"图书情报"领域与"计算机"领域；②通过学科方向组的统计，可以看出"情报学""教育心理学"这些学科方向也被选择得较多。

随后，我们进一步统计了测评人员对推荐结果的评分。在本次 35 份推荐需求中，得到了 112 位推荐用户的评分，各个分值的统计分布结果如图 5 所示：

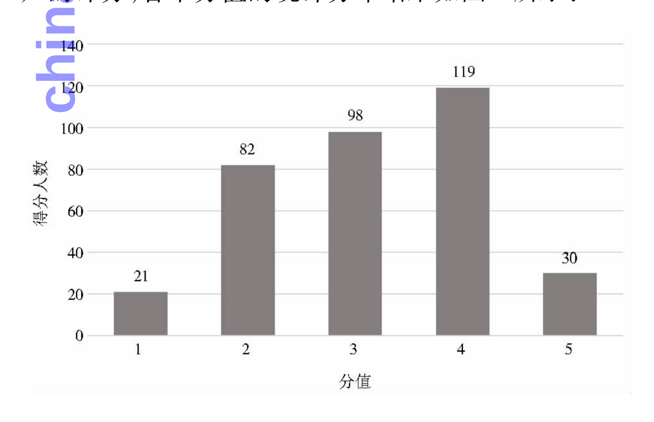


图 5　推荐的跨学科用户得分

通过图 5 可以看出，得到 3 分与 4 分的推荐用户数占据到了整个推荐用户数的 62%，即大部分推荐的跨学科用户得分都在 4 分和 3 分(实际上，平均得分为 3.14 分)，这意味着本文推荐模型具有一定程度的推荐准确度。

5　结论与展望

为了有效融合"媒体"与"社交"进行学术社交媒体上跨学科用户推荐，我们基于前期研究结果设计并实现了一个跨学科用户推荐模型。该模型采用 VSM 来表示跨学科用户，借用欧式距离来计算相似度，并基于推荐指数(结合领域专业度、跨学科距离、学者 PageRank 值)给出推荐用户列表。该模型以科学网上"图书情报""计算机""新闻与传媒""高等教育""生物学"5 个领域的跨学科用户为实验，进行了跨学科用户推荐。为了考核推荐质量，本文邀请了 28 位研究生参与了质量测评，同时通过分析 28 位测试人员的 35 个推荐需求，得到 112 位推荐用户的平均得分为 3.15，这表明该模型具有一定程度的推荐准确度。

本文仅仅围绕科学网上"图书情报""计算机""新闻与传媒""高等教育""生物学"5 个领域内的跨学科用户进行了推荐，未来可以扩展至更多领域、更多跨学科用户的推荐。

参考文献：

[1] HERTZUM M, PEJTERSEN A M. The information-seeking practices of engineers: searching for documents as well as for people [J]. Information processing & management, 2000, 36(5): 761－778.

[2] YIMAM-SEID D, KOBSA A. Expert-finding systems for organizations: problem and domain analysis and the DEMOIR approach

[J]. Journal of organizational computing,2003, 13(1): 1-24.
[3] BARJAK F. The role of the Internet in informal scholarly communication[J]. Journal of the Association for Information Science and Technology,2006, 57(10): 1350-1367.
[4] KIRKUP G. Academic blogging: academic practice and academic identity[J]. London review of education,2010, 8(1): 75-84.
[5] RICHARDS D, TAYLOR M, BUSCH P. Expertise recommendation: a two-way knowledge communication channel[C]//International conference on autonomic and autonomous systems. Guadeloupe:IEEE,2008:35-40.
[6] JENG W, HE D, JIANG J. User participation in an academic social networking service: A survey of open group users on Mendeley [J]. Journal of the Association for Information Science and Technology, 2015, 66(5): 890-904.
[7] DENGLER F, KOSCHMIDER A, OBERWEIS A, et al. Social software for coordination of collaborative process activities[M]. Berlin: Springer,2010.
[8] MEISHAR-TAL H,PIETERSE E. Why do academics use academic social networking Sstes[J]. International review of research in open and distance learning,2017, 18(1): 1-23.
[9] ORTEGA J L. Disciplinary differences in the use of academic social networking sites[J]. Online information review, 2015, 39(4): 520-536.
[10] ELSAYED A M. The use of academic social networks among Arab researchers: a survey[J]. Social science computer review,2016, 34(3): 378-391.
[11] PRIEM J,HEMMINGER B H. Scientometrics 2.0: new metrics of scholarly impact on the social Web[J]. First monday,2010, 15(7):5-9.
[12] GUNN W. Social signals reflect academic impact: what it means when a scholar adds a paper to Mendeley[J]. Information standards quarterly,2013, 25(2): 33-39.
[13] JING L, FENG X, WEI W, et al. ACRec: a co-authorship based random walk model for academic collaboration recommendation [C]//Proceedings of the companion publication of the 23rd international conference on World Wide Web. Republic and Canton of Geneva: ACM,2014:1209-214
[14] ROHANI V A, KASIRUN Z M, KUMAR S, et al. An effective recommender algorithm for cold-start problem in academic social networks[J]. Mathematical problems in engineering,2014(3): 1-12.
[15] SUN O J,JENG W. Groups in academic social networking services - an exploration of their potential as a platform for multi-disciplinary collaboration[C]//IEEE third international conference on privacy, security, risk and trust. Boston: IEEE,2011:545-548..
[16] JIANG J, NI C, HE D, et al. Mendeley group as a new source of interdisciplinarity study:how do disciplines interact on mendeley? [C]// Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries. Indianapolis:ACM,2013:135-138.
[17] WU X,ZHANG C. Finding high-impact interdisciplinary users based on friend discipline distribution in academic social networking sites[J]. Scientometrics,2019, 119(2): 1017-1035.
[18] REICHLING T, KAI S,WULF V. Matching human actors based on their texts:design and evaluation of an instance of the ExpertFinding framework[C]//International ACM siggroup conference on supporting group work. Florida:ACM,2005: 61-70.
[19] 李明, 刘鲁, 王君, 等. 基于模糊文本分类的多知识领域专家推荐方法[J]. 北京航空航天大学学报,2009, 35(10): 1254-1257.
[20] KLEINBERG J M. Authoritative sources in a hyperlinked environment[J]. Journal of the ACM,1999, 46(5): 604-632.
[21] 许云红. 基于网络方法的专家知识推荐[D]. 合肥:中国科学技术大学, 2010.
[22] KAUTZ H, SELMAN B,SHAH M. Referral Web: combining social networks and collaborative filtering[J]. Communications of the ACM,1997, 40(3): 63-65.
[23] 彭兰. 社会化媒体、移动终端、大数据:影响新闻生产的新技术因素[J]. 新闻界,2012(16): 3-8.
[24] SALTON G. A vector space model for automatic indexing[J]. Communications of the ACM,1974, 18(11): 613-620.
[25] 贺颖. 基于科学计量视角的同行评议专家遴选问题研究[D]. 天津:天津大学, 2008.
[26] 高琢玉. 基于多目标决策的专家遴选算法的研究[D]. 长沙:中南大学, 2011.
[27] PORTER A L, COHEN A S, ROESSNER J D, et al. Measuring researcher interdisciplinarity[J]. Scientometrics, 2007, 72(1): 117-147.
[28] 杨良斌, 周秋菊,金碧辉. 基于文献计量的跨学科测度及实证研究[J]. 图书情报工作,2009, 53(10): 87-90.
[29] 和晋飞,房俊民. 一个跨学科性测度指标: 作者专业度[J]. 情报理论与实践,2015, 38(5): 42-45.
[30] BROMHAM L, DINNAGE R,XIA H. Interdisciplinary research has consistently lower funding success[J]. Nature,2016, 534(7609): 684-687.
[31] HELMUS M R, BLAND T J, WILLIAMS C K, et al. Phylogenetic measures of biodiversity[J]. The American naturalist,2007, 169(3): E68-E83.
[32] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: bringing order to the Web[J]. Stanford digital libraries working paper,1998, 9(6): 102-107.
[33] 邱均平,余厚强. 跨学科发文视角下我国图书情报学跨学科研究态势分析[J]. 情报理论与实践,2013, 36(5): 5-10.
[34] 吴小兰,章成志. 基于社交媒体的高影响力跨学科用户发现研究[J]. 情报学报,2017, 36(6): 618-627.

作者贡献说明：
吴小兰：论文构思、起草及实验设计；
章成志：论文修订及实验结果检验设计。

Research on Interdisciplinary User Recommendation Model in Academic Social Media Combining Content and Relations

Wu Xiaolan[1]　Zhang Chengzhi[2]
[1] Department of Internet and New Media, School of Journalism and Communication, Nanjing Normal University, Nanjing 210046
[2] Department of Information Management, School of Economics & Management, Nanjing University of Science and Technology, Nanjing 210094

Abstract: [**Purpose/significance**] With the rapid development of academic social media, when users do interdisciplinary research or seek interdisciplinary cooperation, many scientific research cooperation starts from the acquaintance or attention in social media, so it is very meaningful to research on interdisciplinary user recommendation in academic social media. There are two main types of data in social media: media (represents content) and social (represents relationship). Therefore, this paper recommended interdisciplinary users integrating content and relations. [**Method/process**] After user modeling based on Vector Space Model, this paper calculated user specialization with user content information, measured user's interdisciplinary distance based on relational data, then gave recommendation results combined with PageRank value of user relationship network. [**Result/conclusion**] Taking the science blog as an example, an interdisciplinary user recommendation model in five fields of "Library and Information", "Computer", "News and Media", "Higher Education" and "Biology" been implemented, which has been tested by artificial experiments, and showed that the recommendation results can meet the recommendation requirements to some extent.

Keywords: interdisciplinary users　recommendation model　interdisciplinary distance　academic social networking